US-PAT-NO: 6052686
DOCUMENT-IDENTIFIER: US 6052686 A
TITLE: Database processing using schemas
DATE-ISSUED: April 18, 2000
INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Fernandez; Maria F.	Cranford	NJ	N/A	N/A
Suciu; Dan	Mountainside	NJ	N/A	N/A

US-CL-CURRENT: 707/100; 707/3

ABSTRACT: An apparatus and method for efficiently processing a database. A structure of the database is constructed into a schema which only includes those structures of the database that are known. Desired information to be extracted from the database is specified using path expressions and an automaton models the path expression. A composite automaton is generated based on the automaton and the schema. The composite automaton is pruned and portions of the database corresponding to the pruned automaton is searched to obtain the desired information. Points within the database may be identified to begin searching for the desired information. These points correspond to states of the composite automaton. A hybrid automaton may be formed for each set of possible starting states to determine completeness by simulating the hybrid automaton against the composite automaton.
36 Claims,   10 Drawing figures
Exemplary Claim Number: 1
Number of Drawing Sheets: 7

---------- KWIC ---------

Brief Summary Text - BSTX: The composite automaton may also be analyzed to select subsets of starting states which are below the starting state of the composite automaton. Each subset of starting states may be determined to result in the desired information equivalent to starting at the starting state of the composite automaton. Such determination may be obtained by generating a hybrid automaton and simulating the hybrid automaton against the composite automaton to determine their identity. Where a simulation exists, the subsets of starting states may be compared with each other by using a cost model to identify the subset of starting states that results in the lowest processing costs required for obtaining the desired information.

Detailed Description Text - DETX: The above process may determine that multiple subsets of starting states are equivalent to the starting state DB.S1. Selecting which of the subsets of starting states is the most efficient will require additional cost models associated with each of the subsets of starting states. Costs corresponding to the different subsets of starting states may be compared to select the subset of starting states that is most optimal.